(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 296 403**

(51) Int. Cl.:
G06F 13/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Número de solicitud europea: **99942121 .7**
(86) Fecha de presentación : **11.08.1999**
(87) Número de publicación de la solicitud: **1131725**
(87) Fecha de publicación de la solicitud: **12.09.2001**

(54) Título: **Sistema y procedimiento para analizar archivos de registro de un servidor web.**

(30) Prioridad: **11.08.1998 US 132287**

(45) Fecha de publicación de la mención BOPI:
16.04.2008

(45) Fecha de la publicación del folleto de la patente:
16.04.2008

(73) Titular/es: **WebTrends, Inc.**
851 SW 6th Avenue, Suite 700
Portland, Oregon 97204, US

(72) Inventor/es: **Boyd, William, Glen y**
Shapira, Elijahu

(74) Agente: **Arias Sanz, Juan**

DESCRIPCIÓN

Sistema y procedimiento para analizar archivos de registro de un servidor web.

Antecedentes de la invención

Esta invención se refiere generalmente al análisis de datos de tráfico de servidores web y, más particularmente, a un sistema y a un procedimiento para analizar archivos de registro de un servidor web.

La malla mundial (en lo sucesivo, "web") se está convirtiendo rápidamente en uno de los medios de publicación más importantes de la actualidad. El motivo es simple: los servidores web interconectados por Internet proporcionan el acceso a un público potencialmente mundial con una inversión mínima en tiempo y recursos para construir un sitio web. El servidor web proporciona la posibilidad de obtener y enviar una gran variedad de medios en diversos formatos, incluidos audio, vídeo y el tradicional de texto y gráficos. Y la facilidad con la que se crea un sitio web hace que la posibilidad de llegar a este público mundial sea una realidad para todo tipo de usuarios, desde compañías a empresas de reciente creación, organizaciones e individuos.

A diferencia de otros tipos de medios, los sitios web son interactivos y el servidor web puede recopilar pasivamente información de acceso relativa a cada usuario mediante la observación y el registro de los paquetes de datos de tráfico intercambiados entre el servidor web y el usuario. Se pueden determinar datos importantes relativos a los usuarios de forma directa o por deducción mediante el análisis de los datos de tráfico y el contexto de la petición (o "hit"). Además, los datos de tráfico recopilados a lo largo de un periodo de tiempo pueden proporcionar información estadística, como, por ejemplo, el número de usuarios que visitan el sitio cada día, los países, estados o ciudades desde los que se conectan los usuarios, y el día u hora de la semana más activos. Esta información estadística resulta útil para diseñar estrategias de mercadotecnia o de gestión adaptadas para coincidir mejor con las necesidades aparentes del público. Cada petición está cifrada además con la fecha y la hora del acceso. Debido a que virtualmente toda la información estadística de interés está relacionada con periodos de tiempo, es de vital importancia efectuar un seguimiento preciso de la hora en que se produce cada petición.

Para optimizar el uso de esta información estadística, el análisis del tráfico del servidor web debe realizarse en el momento oportuno. Sin embargo, no es algo extraordinario que un servidor web procese miles de usuarios diariamente. La información de acceso resultante registrada por el servidor web alcanza los megabytes de datos de tráfico. Algunos servidores web generan gigabytes de datos de tráfico diarios. El análisis de los datos de tráfico, incluso para un único día, para identificar tendencias o generar estadísticas, requiere un gran poder de cálculo y mucho tiempo. Además, el tiempo de procesamiento necesario para analizar los datos de tráfico durante varios días, semanas o meses aumenta linealmente a medida que aumenta el marco temporal de interés.

El problema que supone efectuar un análisis del tráfico eficaz y oportuno no es exclusivo de los servidores web. Más bien, el análisis de los datos de tráfico resulta posible siempre que los datos de tráfico puedan observarse y registrarse de modo uniforme, como por ejemplo en una base de datos distribuida, un sistema de cliente-servidor u otro entorno de acceso remoto.

En la patente de EE.UU. nº 5.600.632, a nombre de Schulman, concedida el 4 de febrero de 1997, se describe un procedimiento y un aparato para realizar un seguimiento del rendimiento usando analizadores de red sincronizados. En particular, de acuerdo con el presente documento, se combinan una pluralidad de analizadores de red para dar lugar a una herramienta de análisis del rendimiento de la red para analizar redes complejas y ampliamente distribuidas, estando cada uno de los analizadores sincronizado con los otros para proporcionar uniformidad a los paquetes que atraviesan la red.

Algunos servidores web están tan ocupados, es decir, manejan tanto tráfico, que requieren múltiples servidores para manejar todo el tráfico. Otros usuarios pueden necesitar el uso de múltiples servidores debido al gran tamaño del sitio web. Los sitios críticos, es decir, los que no pueden permitirse estar inactivos debido a algún problema con el servidor, también pueden escoger la opción de desplegar su sitio en múltiples servidores. Estos múltiples servidores se denominan a veces *granja de servidores*. Las granjas de servidores proporcionan un acceso fiable de banda ancha a los sitios web.

Hay varias topologías que se pueden usar en una granja de servidores, pero las más importantes dividen la granja en grupos de servidores. El sitio web está duplicado en cada servidor perteneciente al grupo. Un hardware especial recibe todo el tráfico que entra en el sitio web y distribuye cada *petición* a uno de los servidores. Algunos sistemas proporcionan un preciso equilibrio de cargas, ya que todas las peticiones rotan consecutivamente entre cada uno de los servidores. Pero otros asignan cada *petición* procedente de una nueva fuente a un servidor, y los posteriores accesos al sitio procedentes de esa fuente se dirigen al servidor asignado. Esto se logra asignando un periodo de tiempo predeterminado, por ejemplo 30 minutos, durante el cual todos los futuros accesos desde una misma fuente se consideran parte de una única sesión desde esa fuente. Tal como se describe anteriormente, éste enfoque permite, en cierta medida, el análisis de los archivos de registro, lo cual no es posible usando la técnica de equilibrio de cargas.

Las granjas de servidores, aunque proporcionan equilibrio de cargas y redundancia, presentan problemas en el análisis de los archivos de registro generados por los servidores. Los sistemas para analizar archivos de registro de

servidores web de la técnica anterior pueden manejar múltiples archivos de registro, pero estos archivos se generan de forma consecutiva, es decir, los paquetes de datos contenidos en cada archivo de registro guardan un orden cronológico y los propios archivos de registro corresponden a periodos de tiempo que contienen paquetes de datos pertenecientes a los periodos. En otras palabras, los archivos de registro también se generan de forma consecutiva. No obstante, los archivos de registro contenidos en servidores de una granja de servidores se generan de forma simultánea. Cada archivo de registro cubre o se solapa con el mismo periodo de tiempo. En las granjas de servidores que rotan las *peticiones* entre cada servidor, los programas de análisis de archivos de registro no generan información útil. Existe la posibilidad de emplear soluciones de fuerza bruta, tales como clasificar todos los archivos de registro y crear un nuevo archivo único, o copiar todas las peticiones de cada archivo de registro en una gran base de datos, que pueda clasificar y analizar los datos. Estas soluciones poseen graves inconvenientes: requieren un enorme poder de cálculo y la creación de grandes archivos nuevos, y se realizan sólo cuando los archivos de registro están completos, es decir, no sobre la marcha cuando aún se está completando el archivo de registro.

Las granjas de servidores que asignan a un único usuario peticiones procedentes de una nueva fuente pueden ejecutar en cada servidor programas de análisis de registro de la técnica anterior y sumar los resultados. Sin embargo, esto no resulta completamente preciso y presenta inconvenientes debido a que requiere la generación de distintos informes, cada uno de los cuales debe consultarse o manipularse para obtener información aplicable a toda la granja de servidores.

Por consiguiente, existe una necesidad de disponer de un sistema y un procedimiento para analizar archivos de registro de servidores web que se generen simultáneamente, tales como los generados por una granja de servidores.

Existe también la necesidad de tal sistema y procedimiento que pueda analizar los archivos de registro básicamente en tiempo real.

Existe también además la necesidad de tal sistema que pueda analizar los archivos de registro sin generar grandes archivos nuevos y sin la necesidad de añadir un poder de cálculo sustancial.

También existe la necesidad de tal sistema que pueda analizar los archivos de registro tanto si se han generado de forma simultánea como consecutiva.

Resumen de la invención

La presente invención comprende un procedimiento para analizar archivos de registro que contienen una pluralidad de paquetes de datos en secuencia, que comprende: (a) seleccionar el primer paquete de datos de cada archivo de registro; (b) comparar los paquetes de datos seleccionados; (c) pasar el más antiguo de los paquetes de datos seleccionados a un analizador de archivos de registro; (d) seleccionar el siguiente paquete de datos del archivo de registro en el que se seleccionó el paquete de datos pasado; y (e) repetir las etapas (b) a (d) hasta que se hayan pasado todos los paquetes de datos de los archivos de registro.

Las anteriores y otras características y ventajas de la invención se apreciarán con mayor facilidad tras la siguiente descripción detallada de una forma de realización preferida de la invención, que hace referencia a los dibujos adjuntos.

Breve descripción de los dibujos

La fig. 1 es un diagrama de bloques funcional de un sistema de la técnica anterior para analizar datos de tráfico en un entorno de computación distribuida de acuerdo con la presente invención.

la fig. 2 es un diagrama de flujo de un procedimiento de la técnica anterior para analizar los datos de tráfico en un entorno de computación distribuida de acuerdo con la presente invención, usando el sistema de la fig. 1.

la fig. 3A muestra un formato de la técnica anterior usado en el almacenamiento de una “petición” de datos de tráfico recibida por el servidor de la fig. 1.

la fig. 3B muestra, a modo de ejemplo, una “petición” de datos de tráfico formateados recibida por el servidor de la fig. 1.

la fig. 4 es un diagrama esquemático de una granja de servidores, que incluye múltiples servidores como el que se muestra y describe en la fig. 1.

la fig. 5 es un diagrama esquemático que ilustra el funcionamiento de la granja de servidores de la fig. 4.

la fig. 6 es un diagrama esquemático que ilustra la presente invención incorporada en la granja de servidores de la fig. 4.

la fig. 7 es un diagrama esquemático similar a la fig. 6, pero que ilustra el funcionamiento de la presente invención con archivos de registro consecutivos.

la fig. 8 es un diagrama de flujo de una rutina para incorporar la presente invención.

Descripción detallada de la forma de realización preferida

La fig. 1 es un diagrama de bloques funcional de un sistema de la técnica anterior para analizar datos de tráfico en un entorno de computación distribuida 9. Se describe de forma completa en el documento "WebTrends Installation and User Guide", versión 2.2, octubre de 1996, y en la solicitud de patente de EE.UU. nº 08/801.707, cuyas descripciones se incorporan en la presente memoria descriptiva a modo de referencia. WebTrends es una marca comercial perteneciente a Webtrends Corporation, Portland, Oregón.

Un servidor 10 proporciona a usuarios remotos sitios web y servicios relacionados. A modo de ejemplo, los usuarios remotos pueden acceder al servidor 10 desde un sistema informático remoto 12 interconectado con el servidor 10 a través de una conexión de red 13, como por ejemplo Internet o una red de tipo intranet, una conexión de marcado telefónico (o de punto a punto) 14 o una conexión directa (dedicada) 17. También son posibles otros tipos de conexiones de acceso remoto.

Cada acceso al servidor 10 realizado por un usuario remoto da lugar a una "petición" de datos de tráfico en bruto 11. Más adelante se describe el formato usado al almacenar cada *petición* de datos de tráfico 11 y un ejemplo de *petición* de datos de tráfico 11, haciendo referencia a las figs. 3A y 3B respectivamente. El servidor 10 almacena preferentemente cada petición de datos de tráfico 11 en un archivo de registro 15, aunque se puede usar una base de datos 16 u otra estructura de almacenamiento.

Para analizar los datos de tráfico, el servidor 10 examina cada *petición* de datos de tráfico 11 y almacena la información de acceso obtenida a partir de los datos de tráfico como resultados del análisis 18A a C. Se muestran cinco fuentes de datos de tráfico 11 (sistema remoto 12, conexión de marcado telefónico 14, archivo de registro 15, base de datos 16 y conexión directa 17). También son posibles otras fuentes. Las peticiones de datos de tráfico 11 pueden originarse a partir de cualquier fuente individual o a partir de una combinación de estas fuentes. Mientras el servidor 10 recibe las peticiones de datos de tráfico 11 de forma continua, se almacenan diferentes conjuntos de resultados de los análisis 18A a C para cada periodo de información discreto, denominado fracción de tiempo. Los resultados de los análisis 18A a C se usan para generar resúmenes 19A a C de la información de acceso.

En la forma de realización descrita, el servidor 10 es típicamente un sistema informático basado en Intel Pentium, provisto de un procesador, memoria, interfaces de entrada/salida, un interfaz de red, un dispositivo de almacenamiento secundario y una interfaz de usuario, preferentemente tal como un teclado y una pantalla. El servidor 10 funciona típicamente bajo el control de un sistema operativo Microsoft Windows NT o Unix y ejecuta un software de Microsoft Internet Information Server o NetScape Communications Server. Pentium, Microsoft, Windows, Windows NT, Unix, NetScape y NetScape Communications Server son marcas comerciales de sus respectivos propietarios. No obstante, también son posibles otras configuraciones del servidor 10 con diferente hardware, tal como plataformas compatibles con DOS, Apple Macintosh, Sun Workstation y otras, sistemas operativos, tales como MS-DOS, Unix y otros, y software de red. Apple, Macintosh, Sun y MS-DOS son marcas registradas de sus respectivos propietarios.

La fig. 2 es un diagrama de flujo de un procedimiento 20 para analizar datos de tráfico en un entorno de computación distribuida de acuerdo con la presente invención, usando el sistema de la fig. 1. Su objeto es de el recopilar y resumir continuamente la información de acceso a partir de peticiones de datos de tráfico 11 al tiempo que permite los análisis *ad hoc* cuando se soliciten. El procedimiento 20 consiste en dos rutinas. La información de acceso se recopila a partir de peticiones de datos de tráfico 11 y es resumida por el servidor 10 para dar lugar a los resultados del análisis 18A a C (bloque 21). La información de acceso se analiza por separado para generar los resúmenes 19A a C que identifican las tendencias, estadísticas y otras informaciones (bloque 22). El servidor 10 lleva a cabo de forma continua la recopilación y el resumen de la información de acceso (bloque 21) mientras que el análisis de la información de acceso (bloque 22) se lleva a cabo de forma *ad hoc* mediante el servidor 10 o una estación de trabajo diferente (que no se muestra).

El procedimiento 20 se incorpora preferentemente en forma de programa informático ejecutado por el servidor 10 y materializado en un medio de almacenamiento que comprende un código legible por ordenador. En la forma de realización descrita, el procedimiento 20 está escrito en el lenguaje de programación C, aunque otros lenguajes programación resultan igualmente adecuados. Funciona en un entorno de Microsoft Windows y puede analizar formatos de archivo de registro Common Log File, Combined Log File y de propietario desde servidores web estándar en la industria, tales como los autorizados por NetScape, NCSA, O'Reilly WebSite, Quarterdeck, C-Builder, Microsoft, Oracle, EMWAC, y otros servidores web Windows 3.x, Windows NT 95, Unix y Macintosh. Los resultados de los análisis 18A a C pueden almacenarse en una base de datos 16 de propietario o estándar (que se muestra en la fig. 1), tal como SQL, BTRIEVE, ORACLE, INFORMIX y otras. El procedimiento 20 usa los resultados de los análisis 18A a C de las peticiones de datos de tráfico 11 según se recopilan en el archivo de registro 15 o la base de datos 16 para elaborar resúmenes de actividad, geográficos, demográficos u otros 19A a C, tales como los que figuran en la siguiente tabla 1. También son posibles otros resúmenes 19A a C.

TABLA 1

Perfil de usuario por regiones	Tabla de estadísticas generales
Páginas más solicitadas	Páginas menos solicitadas
Páginas de entrada más frecuentes	Páginas de salida más frecuentes
Páginas de acceso único	Recorridos más frecuentes a través del sitio
Anuncios vistos	Número de clics en anuncios
Anuncios vistos y en los que se ha hecho clic	Archivos más descargados
Organizaciones más activas	Países más activos
Resumen de actividad por día de la semana	Resumen de actividad por día
Resumen de actividad por hora del día	Nivel de resumen de actividad por horas del día
Estadísticas y análisis del servidor web	Errores de cliente
Tipos y tamaños de archivo más descargados	Errores de servidor
Actividad por tipo de organización	Directorios a los que más se accede
Sitios con mayor número de referencias	URLs con mayor número de referencias
Navegadores más usados	Navegadores Netscape
Navegadores Microsoft Explorer	Visitas de arañas
Plataformas más frecuentes	

Además, los resultados de los análisis 18A a C se pueden usar para producir automáticamente informes y resúmenes que incluyan información estadística y gráficos que muestren, a modo de ejemplo, la actividad del usuario por mercado, nivel de interés en páginas o servicios web específicos, qué productos son más populares, si un visitante es de origen local, nacional o internacional, e información similar. En la forma de realización descrita, los resúmenes 19A a C pueden generarse como informes en diversos formatos. Entre estos formatos se incluyen archivos en lenguaje de marcas de hipertexto (HTML) compatibles con la mayor parte de los navegadores web populares, formatos de archivo de propietario para su uso con programas de procesamiento de textos, hoja de cálculo, base de datos y otros, tales como archivos de Microsoft Word, de Microsoft Excel, ASCII y otros formatos diversos. Word y Excel son marcas comerciales de Microsoft Corporation, Redmond, Virginia.

La figura 3A muestra un formato usado en el almacenamiento de una "*petición*" de datos de tráfico en bruto 11 recibida por el servidor de la fig. 1. Una *petición* de datos de tráfico en bruto 11 no se encuentra en el formato que se muestra en la fig. 3A. En lugar de ello, el contenido de cada campo del formato se determina a partir de los paquetes de datos intercambiados entre el servidor 10 y la fuente de la *petición* de datos de tráfico 11 y la información extraída de los paquetes de datos se almacena en un registro de datos usando el formato de la fig. 3A antes de almacenarla en el archivo de registro 15 (que se muestra en la fig. 1) o procesarla.

Cada *petición* de datos de tráfico 11 es una cadena formateada de datos ASCII. El formato está basado en el formato estándar de archivo de registro creado por la National Computer Security Association (NCSA), el formato estándar de registro usado por la mayor parte de servidores web. El formato consiste en los siguientes siete campos:

Nombre del campo	Descripción
Dirección de usuario (30):	Dirección de protocolo de Internet (IP) o nombre de dominio del usuario que accede al sitio.
RFC931 (31):	Campo obsoleto que se suele dejar en blanco, pero que cada vez más usan muchos servidores web para almacenar el nombre de dominio del ordenador principal para archivos de registro con más de una conexión a red (*multi-homed*).
Autenticación de usuario (32):	Intercambia el nombre de usuario si se requiere para el acceso al servidor web.
Fecha/hora (33):	Fecha y hora del acceso y la diferencia horaria con respecto al GMT
Solicitud (34):	Una instrucción GET (una solicitud de página) o POST (un envío de un formulario)
Código de retorno (35):	Estado de retorno de la solicitud que especifica si la transferencia se realizó de forma satisfactoria.
Tamaño de la transferencia (36):	Número de bytes transferidos para la solicitud del archivo, es decir, el tamaño del archivo.

Además, se pueden emplear los siguientes tres campos opcionales:

Nombre del campo	Descripción
Sitio de referencia (37):	URL usada para obtener información del sitio web para efectuar la "*petición*".
Agente (38):	Versión del navegador, incluidos la marca, modelo o número de versión y sistema operativo.
Cookie (39):	Identificador único usado con permiso para identificar a un usuario en particular.

También son posibles otros formatos de peticiones de datos de tráfico 11, incluidos formatos de propietario que contienen campos adicionales, tales como tiempo para transmitir, tipo de funcionamiento del servicio y otros. Además, constantemente se producen modificaciones y adiciones a los formatos de peticiones de datos de tráfico en bruto 11 y un experto en la materia conocería las extensiones requeridas por la presente invención para manejar tales variaciones de los formatos.

La fig. 3B muestra, a modo de ejemplo, una "petición" de datos de tráfico en bruto recibida por el servidor de la fig. 1. El campo de la dirección de usuario 30 es "tarpon.gulf.net", que indica que el usuario procede de un dominio llamado "gulf.net" que reside en una máquina llamada "tarpon". Los campos de RFC931 31 y de autorización de usuario 32 son "-", que indica entradas en blanco. El campo de Fecha/Hora 33 es "12/JAN/1996:20:38:17 +0000", que indica un acceso el 12 de enero de 1996 a las 8:38:17 p.m. GMT. El campo de solicitud 34 es "GET/general.htm http/1.0", que indica que el usuario solicitó la página "general.htm". Los campos de código de retorno 35 y tamaño de transferencia 36 son 200 y 3599, respectivamente, que indican una transferencia satisfactoria de 3599 bytes.

Atendiendo ahora a la fig. 4, una granja de servidores construida de acuerdo con la presente invención se indica generalmente mediante el número 40. En ella se incluyen dos grupos de servidores 42, 44, cada uno de los cuales incluye los servidores 46, 48, 50 y los servidores 50, 52 y 54, respectivamente. Cada uno de los servidores de los grupos 42, 44 es sustancialmente idéntico al servidor 10 de la fig. 1. En la presente forma de realización, el grupo de servidores 42 alberga un primer sitio web, que está duplicado en cada uno de los servidores del grupo, en una única dirección de protocolo de Internet (IP) identificada. Los servidores pertenecientes al grupo 44 albergan un segundo sitio web, que está duplicado en cada uno de los servidores del grupo, en una segunda dirección de IP identificada.

Cada uno de los servidores de los grupos 42, 44 está conectado a través de un cable, como el cable 58, a un redirector 60, El redirector recibe a su vez una entrada procedente de una conexión de red 62, que en la presente forma de realización es una conexión de Internet. El redirector 60 es un dispositivo de hardware de la técnica anterior que

recibe una fuente de peticiones de datos de tráfico -en el presente caso, a través de la conexión 62- y las distribuye a los servidores de los grupos 42, 44.

En la presente puesta en práctica, el redirector 60 distribuye las peticiones de datos de tráfico dentro de cada uno de los grupos 42, 44. En otras palabras, las peticiones de datos de tráfico generados a consecuencia del acceso al sitio web publicado en el grupo 42, se distribuyen entre los servidores 46, 48, 50. Igualmente, las peticiones de datos de tráfico producidas por el acceso al sitio web alojado en el grupo 44 se distribuyen entre los servidores 52, 54, 56. Cisco Systems fabrica un dispositivo adecuado para funcionar como redirector, vendido con el nombre de LocalDirector. Los expertos en la materia apreciarán que otros dispositivos de hardware conocidos pueden realizar la función del redirector 60.

Atendiendo ahora a la fig. 5, cada uno de los archivos de registro 46A, 48B y 50C se almacena en el servidor correspondiente al número usado para indicar el archivo de registro. Estos archivos de registro se generan y se almacenan del modo descrito en relación con el servidor de la fig. 1. En la fig. 5, las peticiones se numeran consecutivamente, de la petición número 1 a la petición número 13 en el orden cronológico en que se generó cada petición de datos de tráfico. En la representación de la fig. 5, aún se sigue completando cada uno de los archivos de registro 46A, 48B, 50C. Es decir, en el archivo de registro 46A, por ejemplo, la petición número 1 es la petición de datos almacenada en primer lugar, y la petición número 5 es la petición de datos almacenada a continuación, almacenándose después, consecutivamente, las peticiones número 6 y 12. Debido a que el archivo de registro 46A no está aún completo y permanece abierto, se pueden almacenar más peticiones consecutivamente tras la petición número 12. El caso es el mismo para los archivos de registro 48B, 50C.

Atendiendo ahora a la fig. 6, en ella se incluye un clasificador 64, que examina las peticiones consecutivamente en cada uno de los archivos de registro y las pasa -en el orden cronológico en que se generó cada petición- a un analizador de archivos de registro 66. El analizador de archivos de registro funciona generalmente de modo descrito en relación con el servidor representado en la fig. 1. Por lo tanto, los resultados del análisis se pasan a los resultados de los análisis 18A a C, también del modo descrito en relación con la fig. 1.

El funcionamiento del clasificador 64 puede entenderse mejor haciendo referencia a la siguiente tabla 2 y al diagrama de flujo representado en la fig. 8.

TABLA 2

Comparar			Pasadas
1	2	4	1
5	2	4	2
5	3	4	3
5	8	4	4
5	8	7	5
6	8	7	6
12	8	7	7
12	8	9	8
12	10	9	9
12	10	11	10
	.		.
	.		.
	.		.
	.		.

En primer lugar, en el bloque 68 de la fig. 8, se selecciona el primer registro recibido en cada archivo de registro 46A, 48B, 50C. Esta selección se ilustra en la tabla 2, línea 1, en la que las peticiones 1, 2 y 4 aparecen en la columna de *Comparar*. En el bloque 70, el clasificador 64 compara cada una de las peticiones 1, 2, y 4 y pasa el registro más antiguo (en el tiempo), concretamente la petición 1 (bloque 72). La rutina determina a continuación, en el bloque 74, si todos los registros en todos los archivos de registro se han seleccionado, comparado y pasado. Si es así, la rutina termina en el bloque 76. Si no, en el bloque 78, la rutina selecciona el siguiente registro del archivo de registro que contiene el registro que se pasó en el bloque 72. En el ejemplo que se está considerando en este momento, el siguiente

registro es la petición número 5 del archivo de registro 46A. A continuación -en referencia a la línea 2 de la tabla 2-, en el bloque 70, se comparan las peticiones 5, 2 y 4, y se pasa la petición 2, al ser la más antigua (en el tiempo) de las tres que se comparan.

Debido a que la rutina funciona en un orden de primero en entrar, primero en salir (FIFO) en cada uno de los archivos de registro, conserva la capacidad de procesamiento mientras los archivos permanecen abiertos y continúan recibiendo más peticiones en orden.

En el ejemplo de la fig. 7, los archivos de registro 80, 82, 84 se ponen en funcionamiento mediante el clasificador 64. Es preciso observar que estos archivos de registro incluyen peticiones que se encuentran en un orden cronológico consecutivo. Lo que es más, cada uno de los archivos de registro se genera en orden cronológico. Así, el archivo de registro 80 representa un periodo de tiempo identificado, que varía entre la hora asociada a la petición 1 y la hora de la petición 4; el archivo de registro 82, entre las horas de las peticiones 5 y 8; y el archivo de registro 84, entre las peticiones 9 y 12. Volviendo a hacer referencia a la fig. 8 y a la tabla 3, que ilustra las comparaciones secuenciales realizadas con los registros de los archivos de registro en la fig. 7, las peticiones 1, 5 y 9 se seleccionan en el bloque 68 y se comparan en el bloque 70. La petición 1, el registro más antiguo, se pasa en el bloque 72 y el siguiente registro en entrar, la petición número 2, se selecciona en el bloque 78. Esta secuencia continúa hasta que se pasan todas las peticiones de la 1 a la 12, pasándose primero en orden las peticiones 1 a 4 del archivo de registro 80, pasándose a continuación en orden las peticiones 5 a 8 del archivo de registro 82, y finalmente las peticiones 9 a 12 en orden del archivo de registro 84.

TABLA 3

Comparar			Pasadas	
1,	5	9	1	1
2,	5	9	2	
3,	5	9	3	
4	5	9	4	
-	5	9	5	
-	6	9	6	
-	7	9	7	
-	8	9	8	
-	-	9	9	
-	-	10	10	

Por consiguiente, la presente invención clasifica correctamente las peticiones de datos de tráfico en archivos generados simultáneamente o generados consecutivamente. Esto resulta ventajoso, ya que obvia la necesidad de rutinas separadas o de configurar un programa dependiendo de si los archivos de registro son consecutivos o simultáneos. Además, la presente invención es capaz de clasificar archivos de registro mientras éstos continúan recibiendo y almacenando nuevas peticiones de datos de tráfico. Este análisis sobre la marcha proporciona a los usuarios datos estadísticos e informes casi en tiempo real.

REIVINDICACIONES

1. Un procedimiento para analizar datos de tráfico generados por una pluralidad de servidores web (10) conectados a través de una red (13) a una pluralidad de dispositivos informáticos (12) que comprende:

(a) generar una pluralidad de peticiones de datos de tráfico (11) para cada acceso por parte de un usuario remoto a uno de los servidores web (10), correspondiendo cada una de dichas peticiones (11) a un paquete de datos intercambiado entre uno de los servidores web (10) y uno de los dispositivos informáticos (12);

(b) asociar las peticiones de datos de tráfico (11) con sus respectivos servidores (10);

caracterizado porque el procedimiento comprende también:

(c) leer una primera petición de datos (11) de cada servidor (10);

(d) comparar la petición de datos de tráfico (11) leída desde un servidor con la petición de datos de tráfico (11) leída desde otro servidor de la pluralidad de servidores web (10);

(e) pasar la petición de datos más antigua (11);

(f) leer la siguiente petición de datos (11) desde el servidor (10) desde el que se leyó la petición de datos (11) pasada;

(g) repetir las etapas (d) a (e) hasta que se hayan leído todos las peticiones de datos (11); y

(h) analizar las peticiones de datos (11) pasadas.

2. El procedimiento de la reivindicación 1, en el que las etapas (b) y (c) se realizan de forma sustancialmente simultánea.

3. El procedimiento de la reivindicación 1, en el que la etapa (b) se realiza antes de la etapa (c).

4. El procedimiento de la reivindicación 1, en el que dichos servidores web (10) se duplican unos a otros.

5. El procedimiento de la reivindicación 1, en el que dichas peticiones de datos de tráfico (11) se generan en orden cronológico y en el que diferentes archivos de registro (15) contienen peticiones de datos correspondientes a peticiones de datos de tráfico (11) generadas en un mismo periodo de tiempo.

6. Un procedimiento para analizar archivos de registro (15) que contienen una pluralidad de peticiones de datos en orden, cada una de las cuales corresponde a una petición de datos de tráfico (11) generada por un servidor web (10), generándose unas peticiones de datos de tráfico (11) para cada acceso de un usuario remoto al servidor web (10), comprendiendo dicho procedimiento:

(a) seleccionar una primera petición de datos (11) en cada archivo de registro (15);

(b) comparar las primeras peticiones de datos (11) seleccionadas de un archivo de registro (15) con la primera petición de datos (11) seleccionada de otro archivo de registro (15);

caracterizado por las etapas de:

(c) pasar la más antigua de las peticiones de datos (11) a un analizador de archivos de registro (66);

(d) seleccionar la siguiente petición de datos (11) del archivo de registro (15) del que se seleccionó la petición de datos (11) pasada; y

(e) repetir las etapas (b) a (d) hasta que se hayan pasado todas las peticiones de datos (11) de los archivos de registro (15).

7. El procedimiento de la reivindicación 6 en el que cada una de dichas peticiones de datos (11) está asociada a una hora única y en la que el último registro de un archivo de registro (15) está asociado a una hora posterior a la del primer registro en otro archivo de registro (15).

8. El procedimiento de la reivindicación 6 en el que cada uno de dichos archivos de registro (15) está asociado a un período de tiempo único y en el que la hora de cada petición de datos (11) está dentro del periodo para su archivo de registro (15).

9. El procedimiento de la reivindicación 6 en el que uno de dichos archivos de registro (15) lo genera un servidor web (10) y en el que otro de dichos archivos de registro (15) lo genera otro servidor web.

10. Un sistema para analizar archivos de registro de servidor web (15) que comprende:

una fuente (12, 14, 15, 16, 17) de peticiones de datos de tráfico (11) generadas por un servidor web (10) para cada acceso por parte de un usuario remoto a los servidores web (10); estando asociada cada una de dichas peticiones de datos (11) a una hora única;

un archivo de registro (15) que contiene las peticiones de datos (11) en orden;

caracterizado por:

un clasificador (64) para clasificar las peticiones de datos (11) de una pluralidad de archivos de registro (15) en orden cronológico; y

un analizador (66) para analizar las peticiones de datos (11) clasificadas.

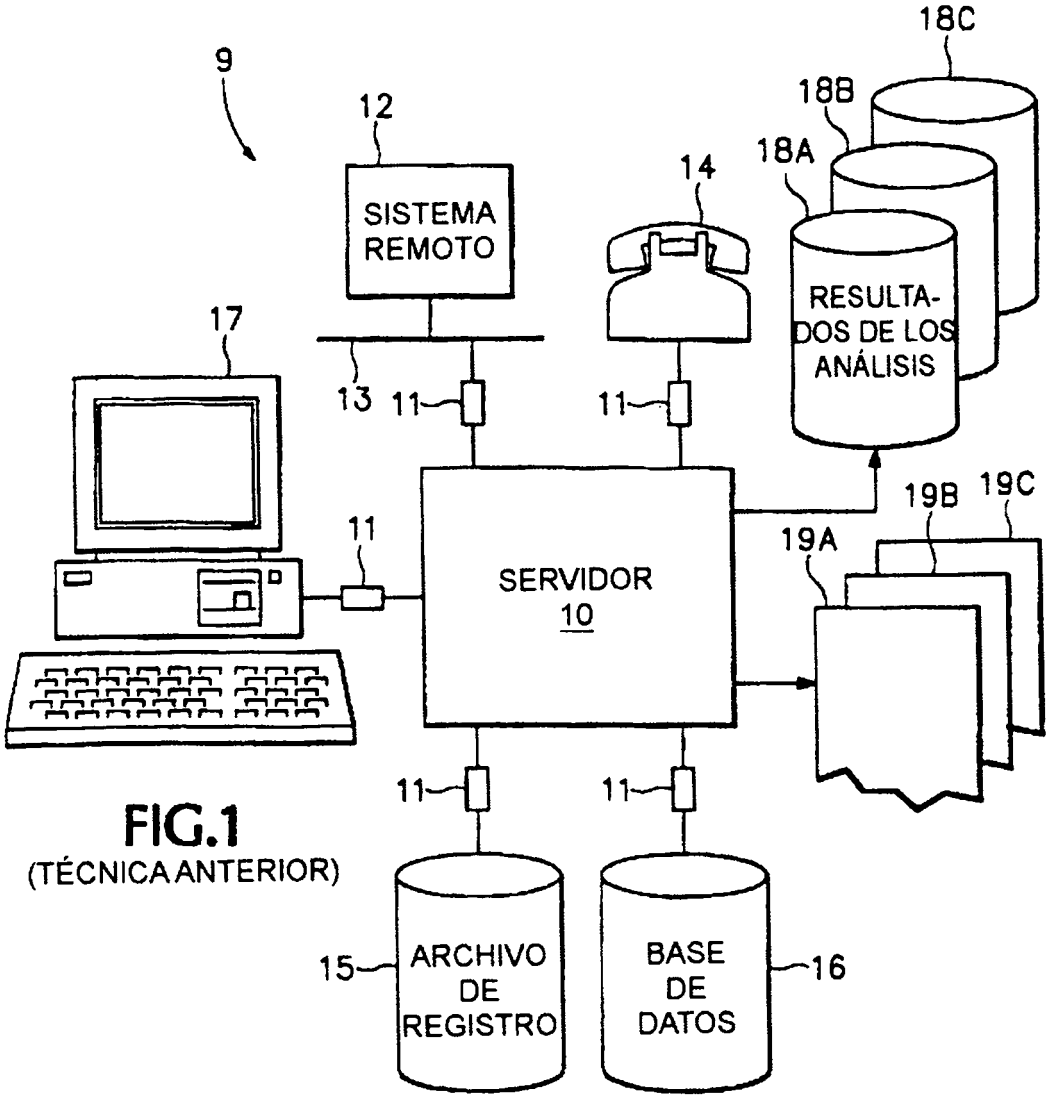


FIG.1
(TÉCNICA ANTERIOR)

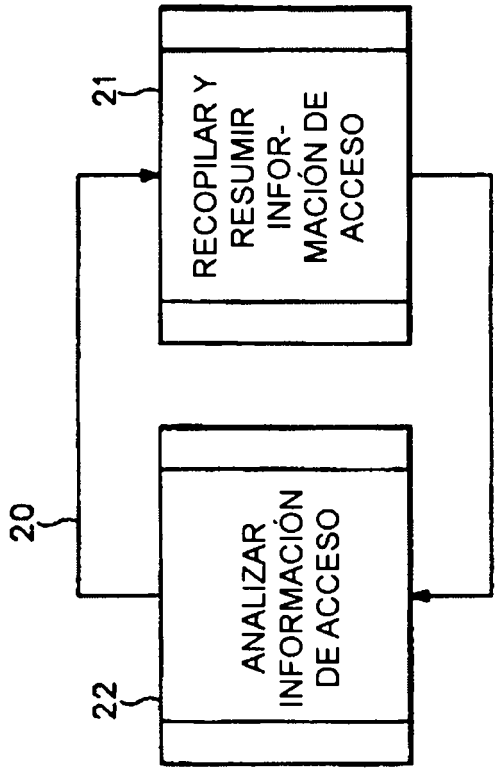


FIG.2
(TÉCNICA ANTERIOR)

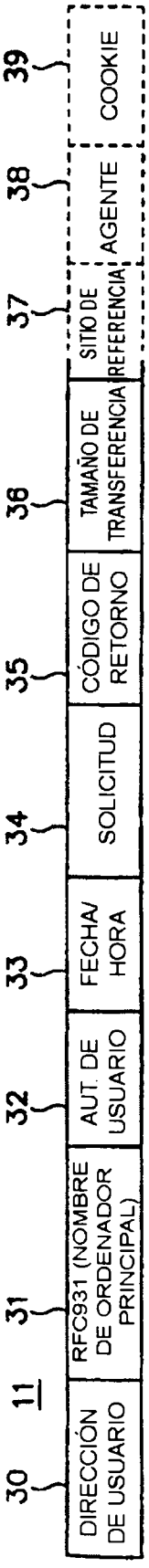


FIG.3A
(TÉCNICA ANTERIOR)

TARPON.GULF.NET--[12/JAN/1996:20:38:17+0000]
"GET/GENERAL.HTM HTTP/1.0"200 3599

FIG.3B
(TÉCNICA ANTERIOR)

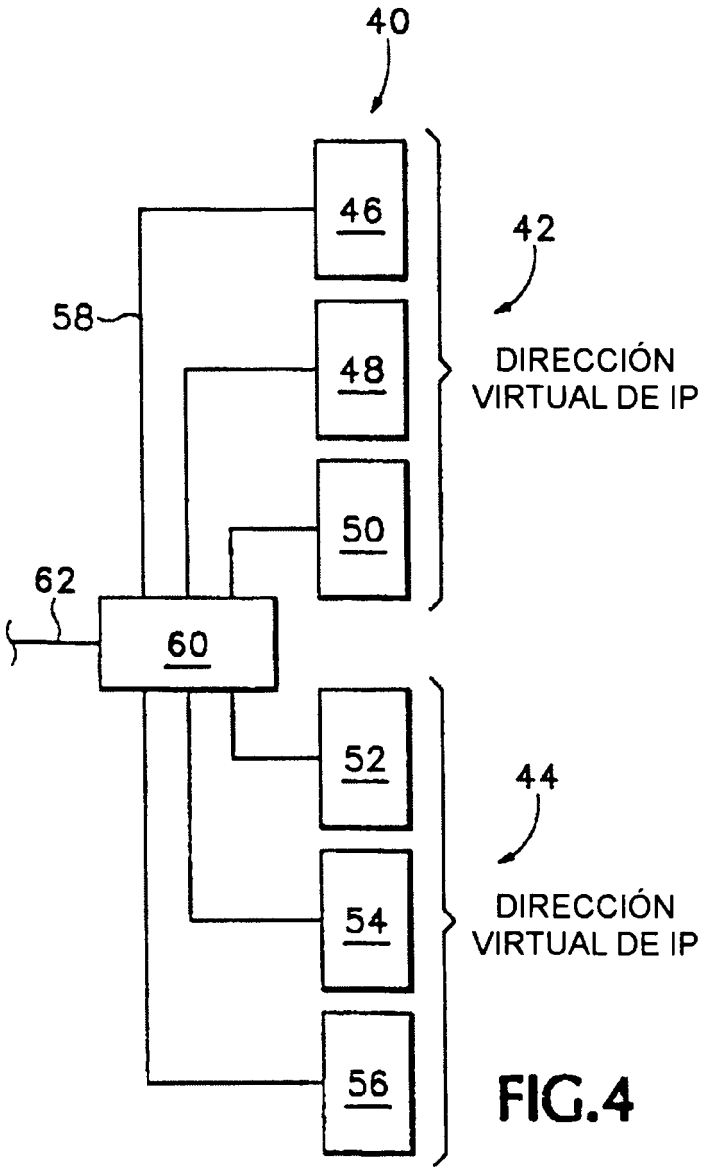
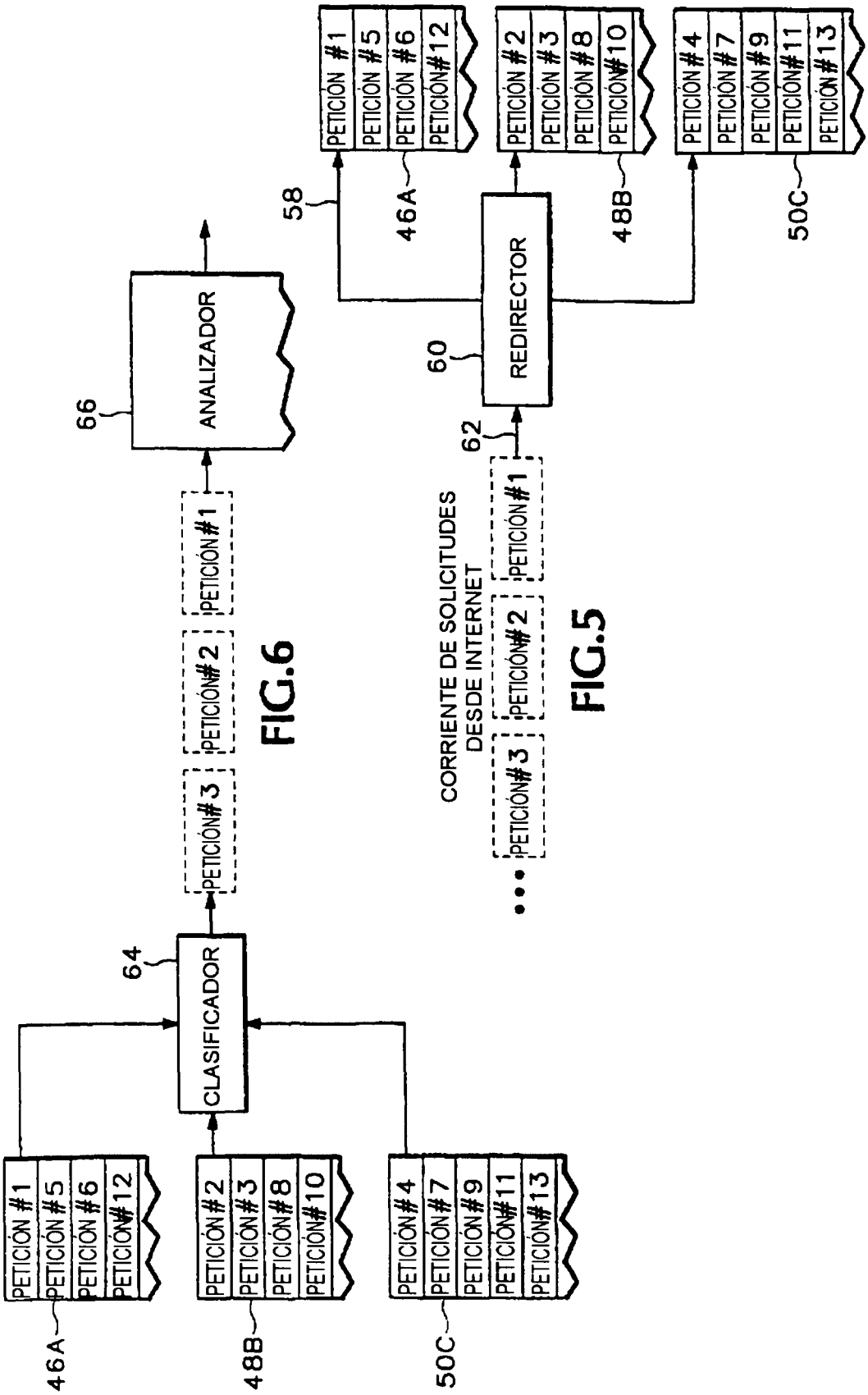


FIG.4


PETICIÓN #1
PETICIÓN #5
PETICIÓN #6
PETICIÓN#12
46A
PETICIÓN#2
PETICIÓN#3
PETICIÓN#8
PETICIÓN#10
48B
PETICIÓN#4
PETICIÓN#7
PETICIÓN#9
PETICIÓN#11
PETICIÓN#13
50C
CLASIFICADOR
64
PETICIÓN#3
PETICIÓN#2
PETICIÓN#1
ANALIZADOR
66
FIG.6
CORRIENTE DE SOLICITUDES DESDE INTERNET
PETICIÓN#3
PETICIÓN#2
PETICIÓN#1
62
REDIRECTOR
60
58
FIG.5

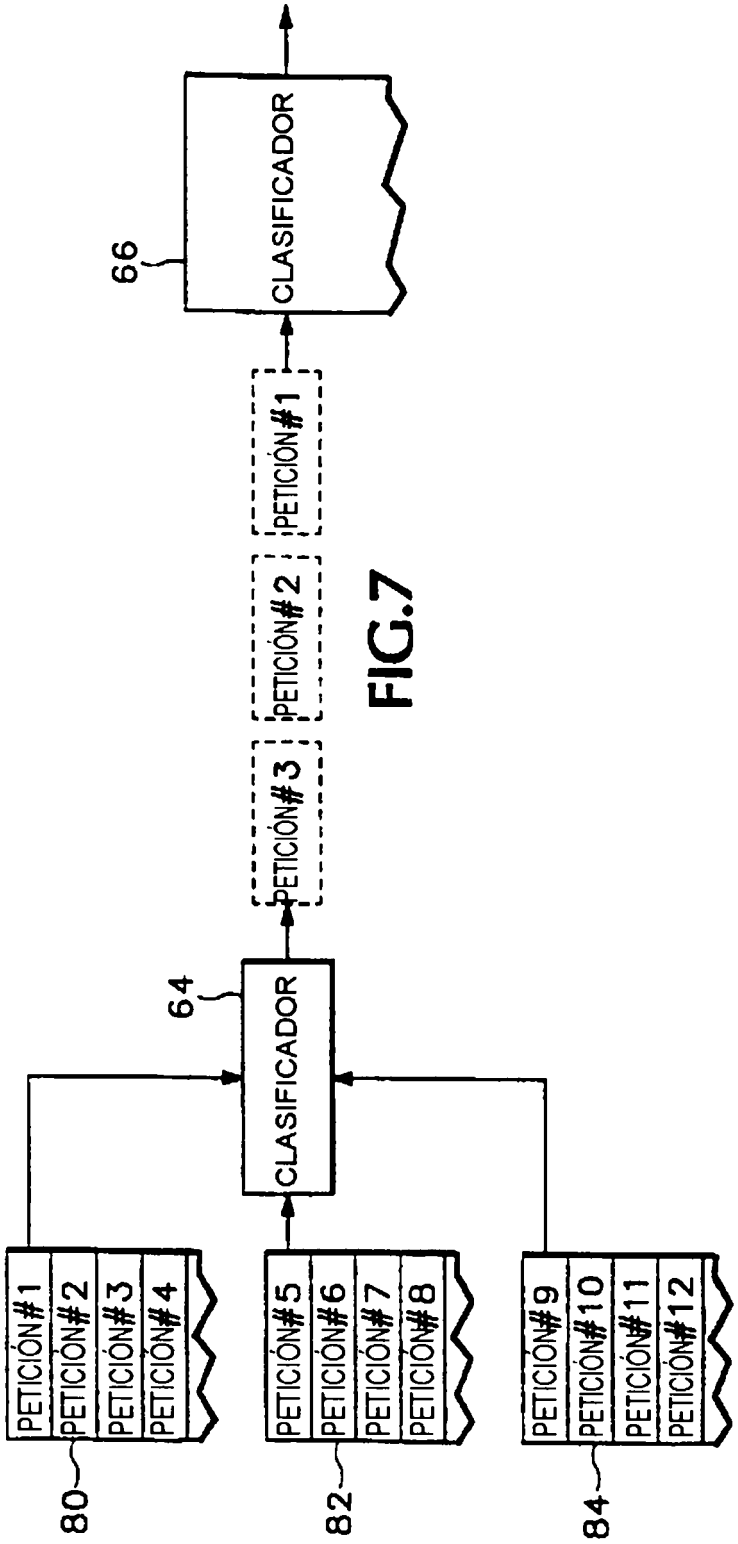

CLASIFICADOR
66
PETICIÓN#1
PETICIÓN#2
PETICIÓN#3
CLASIFICADOR
64
PETICIÓN#1
PETICIÓN#2
PETICIÓN#3
PETICIÓN#4
80
PETICIÓN#5
PETICIÓN#6
PETICIÓN#7
PETICIÓN#8
82
PETICIÓN#9
PETICIÓN#10
PETICIÓN#11
PETICIÓN#12
84

FIG.7

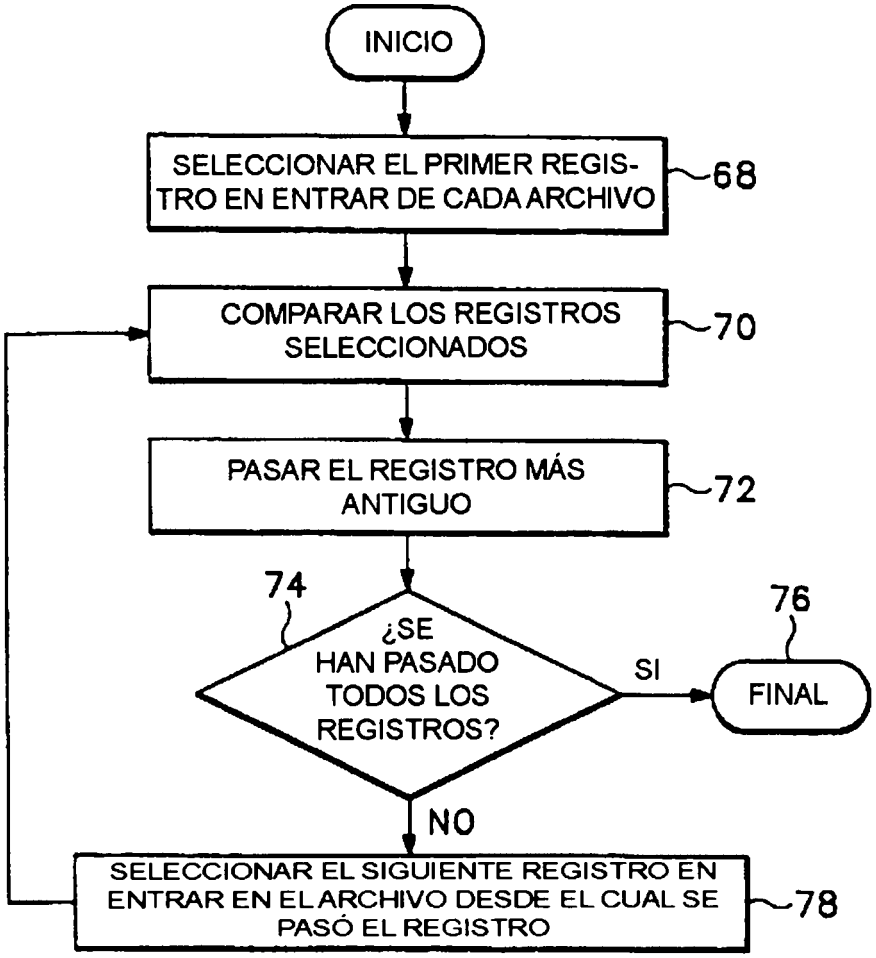

INICIO
SELECCIONAR EL PRIMER REGIS-TRO EN ENTRAR DE CADA ARCHIVO
68
COMPARAR LOS REGISTROS SELECCIONADOS
70
PASAR EL REGISTRO MÁS ANTIGUO
72
74
¿SE HAN PASADO TODOS LOS REGISTROS?
SI
76
FINAL
NO
SELECCIONAR EL SIGUIENTE REGISTRO EN ENTRAR EN EL ARCHIVO DESDE EL CUAL SE PASÓ EL REGISTRO
78

FIG.8